

O63AIntO

1 UNITED STATES DISTRICT COURT
2 SOUTHERN DISTRICT OF NEW YORK

-----x

3 THE INTERCEPT MEDIA, INC.,

4 Plaintiff,

5 v.

24 Civ. 1515 (JSR)

6 OPEN AI, INC., MICROSOFT, et
7 al.,

Oral Argument

8 Defendants.

9 -----x

New York, N.Y.
June 3, 2024
10:10 a.m.

11 Before:

12 HON. JED S. RAKOFF,

District Judge

13 APPEARANCES

14 LOEVY & LOEVY

Attorneys for Plaintiff

15 BY: STEPHEN S. MATCH
16 MATTHEW TOPIC

17 MORRISON & FOERSTER LLP

Attorneys for Defendant OpenAI

18 BY: JOSEPH C. GRATZ
19 ERIC NIKOLAIDES

20 LATHAM & WATKINS LLP

Attorneys for Defendant OpenAI

21 BY: YIJUN ZHONG
22 ALLISON L. STILLMAN

23 ORRICK, HERRINGTON & SUTCLIFFE LLP

Attorneys for Defendant Microsoft

24 BY: ANNETTE L. HURST
CHRISTOPHER CARIELLO
LISA T. SIMPSON

25 Also Present:
David Bralow

O63AIntO

(Case called)

MR. MATCH: Good morning. Stephen Match for The Intercept. With me I have Matt Topic, and David Bralow, general counsel of The Intercept.

THE COURT: Good morning.

MR. GRATZ: Good morning, your Honor. Joe Gratz for Morrison Foerster for OpenAI. With me are Eric Nikolaides from Morrison Foerster, and Alli Stillman and Yijun Zhong from the Latham Firm.

THE COURT: Good morning.

MS. HURST: Good morning, your Honor. Annette Hurst from Orrick, Herrington & Sutcliffe for Microsoft. And with me this morning are my colleagues Lisa Simpson and Chris Cariello.

THE COURT: Good morning. Were here on the motion to dismiss. Let me hear first from moving counsel.

MR. GRATZ: Thank you, your Honor.

Your Honor, the plaintiffs here are apparently not able to get ChatGPT to output their material. And you would think that would make them happy. It certainly makes us happy. That's not what ChatGPT is for, but it causes a problem for their claim because it means they have not been injured in a concrete way.

THE COURT: Let's take a hypothetical. So supposing you're a reporter for a Chicago newspaper called the *Chicago Examiner*. Your name is Hildy Johnson. And you're all too

O63AIntO

1 young to know what I'm talking about.

2 MR. GRATZ: *The Front Page*, your Honor?

3 THE COURT: You got it. And you run an exclusive
4 interview with accused murderer Earl Williams.

5 MR. GRATZ: Indeed.

6 THE COURT: Fully copyright. And now fast forward to
7 the present, and it's picked up, along with a million other
8 things, by ChatGPT. And then you are a consumer and you go on
9 and you say, what would Earl Williams have to say if he were
10 asked to give his own account of why he committed the alleged
11 murder. And out pops word for word the story that we've just
12 referred to, but without any copyright information, any
13 indication it came from the *Chicago Examiner*, etc.

14 If that were to occur, would you agree that would be a
15 violation?

16 MR. GRATZ: So, your Honor, I want to add, I want to
17 add a fact to your hypothetical, which is the material that
18 popped out is copyright expression, copyrighted by the
19 newspaper. Right?

20 THE COURT: Right.

21 MR. GRATZ: And I think in that case that might well
22 be a problem. I think that might well be a problem for under a
23 number of theories. It might be a copyright infringement
24 problem, to which there might be a fair use question attached.
25 There might be a number of issues. But there might be, if

O63AIntO

1 there were all the elements there, if there were reason to
2 believe that that infringement was concealed beforehand, that
3 might well be a 1202 problem if you have the identical thing
4 coming out without -- and the only difference was it didn't
5 have the CMI attached.

6 THE COURT: So your point, as I take it, is that they
7 haven't even come close to something like that, whereas,
8 theoretically, they could have inputted something into your
9 chat, and if out came their stories, they would have a claim.

10 MR. GRATZ: Or at least they would be much closer to
11 having a claim, your Honor, yes.

12 THE COURT: Okay. So I know we want to go in order
13 here, but let me ask you to sit down for a minute, and we'll
14 come back to you very shortly.

15 MR. GRATZ: Thank you, your Honor.

16 THE COURT: But let me ask plaintiff's counsel, why
17 isn't that the problem?

18 MR. MATCH: Good morning, your Honor. It's not a
19 problem because 1202(b)(1) and 1202(b)(3) don't require there
20 to be an actual infringement or an actual concealment of an
21 infringement. If we're talking about what we need to say to
22 state a claim as opposed to standing, which is where I think
23 Mr. Gratz was starting here, 1202(b)(1) requires that there be
24 a possible concealment of infringement in the future. And
25 we --

O63AIntO

1 THE COURT: How do you know for starters, because I'm
2 not sure I saw it in your complaint other than in conclusory
3 fashion, that they even copied your stuff?

4 MR. MATCH: So we've alleged that there are recreated
5 approximations of their training sets and that those training
6 sets contain --

7 THE COURT: Recreated proximations?

8 MR. MATCH: Approximations, yes.

9 THE COURT: What the heck are they?

10 MR. MATCH: I have to go outside the complaint to
11 describe this, but they are --

12 THE COURT: Well --

13 MR. MATCH: If you want me to stay within -- we
14 haven't described them in detail in the complaint.

15 THE COURT: So what am I to make of something you
16 haven't specified?

17 MR. MATCH: I think we've alleged plausibly that there
18 are these approximations created by third parties of their
19 training sets that contain our articles in them.

20 THE COURT: You may have alleged that, but if you
21 haven't alleged that -- where do you allege that the
22 approximate ones show that your stuff was copied?

23 MR. MATCH: I'm sorry, that our stuff is copyrighted?

24 THE COURT: No. It has been --

25 MR. MATCH: Copied.

O63AIntO

1 THE COURT: -- copied by them.

2 MR. MATCH: That would be paragraph 38.

3 THE COURT: Hold on a minute. Paragraph 38.

4 "Various sources," p.s. unnamed, "have recreated
5 approximations of the Common Crawl and WebText training sets
6 based on publicly available information discussing the
7 methodologies used to create them. Those sources have made
8 these recreated data sets, or instructions on how to derive
9 them, available on the internet. Thousands of plaintiff's
10 works are contained in the recreated versions of these data
11 sets without the author, title, copyright notice, and terms of
12 use infringement found in plaintiff's original publications."

13 So if I were to give you leave to amend, what would
14 you add to that paragraph?

15 MR. MATCH: So if you were to give us leave to amend,
16 we would explain with more detail about what these
17 approximations were and how they were derived. If you want I
18 can say it now without --

19 THE COURT: Absolutely.

20 MR. MATCH: Sure. There are a couple of different
21 ones. There's one called C4. C4 is a training set that was
22 used for Google's own AI model, another large language model,
23 kind of like ChatGPT. It was created from the same overall
24 corpus that some of OpenAI's training sets were created from.
25 That's called Common Crawl, which we've identified in this

O63AIntO

1 paragraph.

2 And we do allege this in the complaint that we
3 employed a data scientist to analyze the C4 data set. And the
4 C4 data set contains 2,753 Intercept articles. And he was able
5 to write a script that has identified for each article exactly
6 what type of copyright management information was removed from
7 it. And in --

8 THE COURT: So let me stop you there.

9 MR. MATCH: Yes.

10 THE COURT: So assuming for the sake of argument that
11 that's what was done by the defendants here, the equivalent of
12 what you've just described. Are you saying that their mere
13 input of those into their database is a violation, or do you
14 contend that the result was that the output would be a
15 violation?

16 MR. MATCH: So, both. So the copying into the
17 database, sort of the initial copying into the database, if it
18 could be like an infringement of copyright under Section 106,
19 we haven't brought an infringement claim here because our works
20 aren't registered with the Copyright Office. That's not
21 required for a 1202 claim, which is the claim we're bringing
22 here. So the copying of the articles and the removal of the
23 copyright management information is a violation of 1202(b)(1).

24 THE COURT: Why is your stuff not copyright?

25 MR. MATCH: Well, we own copyright to it.

O63AIntO

1 THE COURT: I understand that. But I mean why is it
2 not registered?

3 MR. MATCH: It hasn't been registered because unlike
4 for print publications, there's no sort of -- it's economically
5 just very expensive to register copyrights in online only
6 publications because there's no ability to bulk register them.
7 So that's why, I mean, I think very few, if any, online only
8 establishers have registered their online news articles with a
9 Copyright Office. Unlike, you know, more legacy media that
10 also has a print version that can bulk register their works.
11 So that's why it's not registered.

12 THE COURT: So I cut you off sort of midstream. You
13 were saying it's both the violation under both the input and
14 the output, and what about the output aspect?

15 MR. MATCH: So, the output, if our works are conveyed
16 to the public, without copyright management information, that
17 would be a violation.

18 THE COURT: Yes. But supposing like it's a paragraph
19 here or a paragraph there, still a violation?

20 MR. MATCH: Well, I mean there's -- this isn't an
21 issue that really has been briefed in a lot of detail. There
22 has been cases holding that excerpts produced without CMI can
23 constitute a violation. Now, I recognize that there has to be
24 a somewhat close to an identical match between the original
25 article and the copied version without CMI under (b)(3). But

O63AIntO

1 if we're talking about the scienter element common to both, you
2 know, knowledge that an infringement will be concealed, which
3 relates to the removal claim as well, that doesn't have to be
4 an identical copy because infringement itself doesn't require
5 anything like an identical reproduction. It could be a large
6 amount of text. I mean, there you have to get into fair use.
7 But for the scienter element, concealment of an infringement,
8 that doesn't have to be an identical copy.

9 THE COURT: So what stopped you from just putting
10 questions into the defendant's product and seeing if what came
11 out was your stuff?

12 MR. MATCH: So, this is not the first of these cases
13 that's been filed against the defendants. In the *New York*
14 *Times* case, they had regurgitations of their works. They put
15 them in the complaint, and they were accused by OpenAI of
16 hacking their products.

17 So we don't believe that we had to do that in this
18 case, both for purposes of (b)(1) -- I'm sorry, 12(b)(6) motion
19 and purposes of standing. If the Court concludes that that's
20 required, I would request permission to replead those facts and
21 try to do that. But we haven't done that because of the
22 accusations that OpenAI made against *New York Times* in its
23 case.

24 THE COURT: Okay. Let's go back to defense counsel.
25 We'll come back to plaintiff's counsel in a minute.

O63AIntO

1 So, first of all, anything you wanted to say just in
2 response to what was just said by plaintiff's counsel?

3 MR. GRATZ: Only, your Honor, that we heard a lot of
4 "if" and a lot of "it's possible." And the *CoreLogic* case from
5 Judge Berzon addresses how you think about those issues in the
6 context of this statute in way that I think sort of resolves
7 why this complaint doesn't.

8 THE COURT: So am I right that your training sets are
9 kept secret?

10 MR. GRATZ: That is correct. We do not make all of
11 our training sets public.

12 THE COURT: And so how is any plaintiff who believes
13 you have violated the relevant statutes going to be able to
14 plead, if other than by using say analogous training sets or
15 other expert material or whatever, as they allege they've done
16 here?

17 MR. GRATZ: The answer is if it never comes out, none
18 of them have a claim. And they can see whether or not it comes
19 out.

20 THE COURT: And they can see by putting in their
21 own --

22 MR. GRATZ: They can, they can --

23 THE COURT: So why shouldn't I give him, give the
24 plaintiffs -- plaintiff, who I guess is her. Right? Plaintiff
25 is -- no, it's it.

O63AIntO

1 MR. GRATZ: It.

2 THE COURT: Or these days we would say they. Why
3 shouldn't I give the plaintiff an opportunity to replead to see
4 if they can do that?

5 MR. GRATZ: Well, I think the first thing I would say
6 is they've made no proffer that they would be able to. What
7 they've said is well, you know -- and they have not told us
8 whether or not they tried and were unable to.

9 What they have said is, well, *The Times* did what they
10 did and you said what they did broke the rules and was
11 inappropriate, and so we didn't try and do anything. They
12 should have tried to sweet talk ChatGPT into doing something it
13 wasn't designed to do to see whether they had a claim here. We
14 don't think they're going to be able to and they've made no
15 proffer that they're able to. And, you know, we have -- you
16 can get to this from your phone. This isn't a hidden or
17 difficult thing.

18 And so I think the answer is they haven't made a
19 proffer of what would happen if they got leave to do that. And
20 I think, in fact, they just haven't been -- we heard they have
21 a data scientist. They haven't been able to do it.

22 THE COURT: So I come back, since your argument in
23 substantial part is they haven't done that, they therefore,
24 haven't alleged any actual injury, therefore they both lack
25 standing and can't make a claim. And given the thrust of the

O63AIntO

1 federal rules for free leave to amend, why shouldn't I give
2 them say two weeks to give it a shot?

3 MR. GRATZ: We think that is within your power. We
4 don't think they've made a proffer that like necessarily
5 justifies it, but we recognize it's within your Honor's power
6 to do so, and we can be back here arguing about whatever other
7 issues arise out of whatever they find or don't find.

8 THE COURT: And think how much more you could bill
9 your clients. We'll disregard that.

10 Let's go back to plaintiff's counsel. Assuming for
11 the sake of argument I give you two weeks to amend in the
12 fashion I've just indicated and you've tried, fail. You can't
13 get anything that's remotely your stories out of the stuff, is
14 that the end of your lawsuit?

15 MR. MATCH: No. And I want to take that both from the
16 perspective of standing and from the perspective of the
17 elements of the claim. So on standing, the issue is really
18 whether we've alleged a concrete injury and whether
19 dissemination of our works is required for that.

20 The *TransUnion* framework is the applicable framework
21 here, which requires us to identify a historical analogue to
22 our injury. I think we don't need to identify an exact
23 duplicate. And I think the Second Circuit's decision in *Saba*
24 *Cap. v. Nuveen* provides a helpful starting point for how close
25 that analogy has to be. It found sufficiently analogous an

O63AIntO

1 analogy between trespass to channels and interference with
2 voting rights in an investment fund, solely because they both
3 involved something analogous to a property-based injury. And
4 that's essentially what we're identifying as the relevant
5 injury for purposes of standing, which is a property-based
6 injury analogous to copyright infringement.

7 So infringements don't require dissemination. We
8 briefed this, and they haven't disputed it. So if copyright
9 infringement is the relevant analogy, then I think we've
10 satisfied the concreteness requirement even if we haven't
11 identified any dissemination. Why is it a close analogy? It's
12 a close analogy because I think in typical cases, violations
13 of --

14 THE COURT: How are you injured, assuming your
15 premise, which is that they're copying, period, regardless of
16 whether they then disseminated is a violation; how are you
17 injured by that fact?

18 MR. MATCH: The injury is the interference with our
19 property. And that has been recognized as an injury since --

20 THE COURT: Well, I know those cases in the copyright
21 context. What's your damages?

22 MR. MATCH: So in this case we're seeking statutory
23 damages primarily, which is provided by the DMCA. It's also
24 provided by the Copyright Act, which Congress did because it
25 recognized that actual damages can be very difficult to prove

O63AIntO

1 in these sorts of cases. So that's primarily what we're
2 seeking here.

3 THE COURT: Okay. Let me go back. I'm sorry to go
4 back and forth.

5 MR. MATCH: No problem.

6 THE COURT: It's more helpful for me to do it this
7 way.

8 MR. GRATZ: And, your Honor, Ms. Hurst will address
9 this line.

10 THE COURT: Okay. Very good. Thank you.

11 So supposing, hypothetically, an unquestionably
12 copyrighted work is fed into your database but never
13 disseminated, why isn't that nevertheless a violation?

14 MS. HURST: Your Honor, it's clear that the
15 attribution right is not part of the historical tradition of
16 copyright in the United States. And we can see this, your
17 Honor, because the right of attribution did not come into the
18 Copyright Act until the 1990 enactment of Section 106A in the
19 Visual Artists Rights Act.

20 And for all the years prior to that, your Honor, not
21 only was the right of attribution not part of the U.S.
22 Copyright Act, but the U.S. had refused on a number of
23 occasions to exceed to The Berne Convention and the provisions
24 that included the right of attribution and the right of
25 integrity.

O63AInto

1 Your Honor, the Second Circuit describes the fact that
2 moral rights, *droit moral* from the French tradition, are not
3 part of the Copyright Act in the *Gilliam* case. And there's
4 also an extensive tracing of this history in a report issued by
5 the Copyright Office in 2016, your Honor. And let me just give
6 the Court --

7 THE COURT: 2016, oh, yeah, just shortly before my
8 birth.

9 MS. HURST: Apologies, your Honor. It's 2019
10 actually, your Honor, and that report is available on the
11 Copyright Office website. It's a report from the register
12 Karyn Temple entitled *Authors, Attribution, and Integrity:*
13 *Examining Moral Rights in the United States.*

14 And, your Honor, it's really clear that the right of
15 attribution, which is what CMI protects, just isn't part of the
16 historical tradition of protection of economic rights for
17 copyright. And indeed, it really excludes the possibility of
18 treating that as closely analogous to a copyright interest
19 because it has never been so. And even now, your Honor,
20 there's only a very limited right of attribution in Section
21 106A.

22 So the CMI protects this attribution, a voluntary
23 attribution that people can apply by putting CMI on it, but it
24 is that interest in personality or the right to be associated
25 with the work that is protected. And that is what has to be

O63AIntO

1 harmed here, your Honor, in order for them to have a
2 sufficiently concrete injury under 12(b)(1).

3 Without dissemination, your Honor, that interest in
4 personality simply cannot be harmed in the way that is required
5 by *TransUnion*. And those were really the facts in *TransUnion*,
6 your Honor, where the Court made a distinction between those in
7 the class who had been able to show dissemination, and those in
8 the class who had not.

9 So, your Honor, because moral rights are not part of
10 the U.S. copyright tradition, the right of attribution is
11 limited to dissemination. Without outputs, your Honor, they
12 cannot plead the harm that is required to meet the 12(b)(1)
13 standard. And, your Honor, I would just note that the words
14 "harm" or "injury" do not appear anywhere in this complaint.
15 There is presently no theory of harm in this complaint.

16 THE COURT: What do you say to the argument that
17 Congress in this area has traditionally recognized that copying
18 is itself a harm and that's why they put in statutory damages
19 because often you can't prove any other harm?

20 MS. HURST: That is correct, your Honor. And the
21 answer to that is to bring a claim under the Copyright Act with
22 the intended restrictions and limitations on that claim, the
23 Section 1202 claim was not meant to supplant those restrictions
24 and limitations. And indeed, the scheme there requires injury
25 as a statutory matter in Section 1203 that is not part of

O63AIntO

1 Section 501, in the main part of the Copyright Act, your Honor.

2 So there's a clear distinction and congress recognized
3 that limitation as part of the DMCA, whereas it has not
4 previously recognized that limitation as part of Section 501 or
5 Section 504.

6 THE COURT: Okay. Let me hear from plaintiffs on this
7 issue.

8 MR. MATCH: So the DMCA is not fundamentally about
9 attribution. The strategy that Ms. Hurst is employing here is
10 to suggest that the appropriate analogy that the Court needs to
11 consider here is between the DMCA on the one hand and a
12 nonexistent attribution right on the other. According to
13 *TransUnion*, though, the question is whether the plaintiff has
14 identified a close historical analogue.

15 So there was a question about whether we get to pick
16 our analogy. We do under *TransUnion*. Of course the Court has
17 to decide whether it's close enough. But I just want to point
18 out why attribution is not a particularly close analogy here.
19 There's a number of reasons.

20 Let's start with the report that Ms. Hurst mentioned
21 from the Copyright Office. I agree that that report says that
22 attribution is not part of the sort of tradition of copyright
23 in this country. That report though -- I mean as she pointed
24 out, Section 106A, the Visual Artists Rights Act, expressly
25 mentions attribution in the text.

O63AInto

1 The DMCA says nothing about attribution. You would
2 think that if the DMCA were about attribution, it would require
3 attribution. But it doesn't. The DMCA does not require
4 anybody to put any CMI on any work. It allows you, if you want
5 to put CMI in the work, you can do that, but you don't have to
6 put the author on. You can give plenty of information that
7 qualifies as copyright management information that is not
8 attributed, such as the title of the work, that doesn't
9 attribute the work to anybody.

10 The page 98 of the report that Ms. Hurst mentioned,
11 contains a recommendation from the Copyright Office that
12 Congress amend Section 1202 to add a new section that would be
13 identical to the present version of the statute. Except that
14 it replaces the current scienter element with one that says,
15 you know, with knowledge that the work wouldn't be attributed.

16 And attribution, I think if that were really the point
17 of the DMCA would -- it's hard to see how Section 1202(b)(1)
18 would even further that purpose at all given that we have
19 Section 1202(b)(3), which requires distribution. (b)(1)
20 doesn't. It can be satisfied by removal.

21 So the point is the defendants want the Court to sort
22 of ask whether attribution is the right analogy and find that
23 it's not satisfied because there's no dissemination. But the
24 point is copyright infringement is a close analogy and it
25 doesn't require dissemination. I would like to point out that

O63AIntO

1 infringement will often be the way by which someone violates
2 1202(b)(1) and 1202(b)(3). So one way to violate (b)(1) is by
3 creating a copy of a work that has CMI on it, making a copy of
4 the work, and not including CMI on the copy. A citation for
5 that is *Associated Press v. All Headline News Corp*, which we
6 cite in our brief.

7 So that will violate (b)(1) assuming the scienter
8 elements are met. But it will also violate at least prima
9 facie the exclusive right to reproduction under Section 106.
10 Ditto for (b)(3). (b)(3) requires distribution of a copyright
11 protected work, which is also one of the exclusive rights under
12 Section 106. In fact, I think the (b)(3) claim, which we
13 haven't discussed a lot today, even satisfies their theory of
14 standing at least as requiring dissemination, because by
15 definition a distribution is a dissemination of the work.

16 So the point I want to drive home is that copyright
17 infringement is a close analogy that DMCA, they both involve
18 tampering with copyright protected works. Even on a (b)(1)
19 claim, even if it's not done by actual copyright infringement,
20 it's at least close to it because you're taking off information
21 that's conveyed very close to the work. In the case of a news
22 article, maybe you're removing the headline, which is what we
23 alleged. The headline, although it's not copyrighted, it
24 conveys meaning that is central to the meaning of the actual
25 work. I think common experience shows that when you read a

O63AIntO

1 news article, you're sort of colored by the headlines. So
2 that's analogous to copyright infringement. It doesn't have to
3 be infringement. *TransUnion* was very clear that it doesn't
4 have to be exact duplicate. It's close and that's what we need
5 to allege standing.

6 THE COURT: All right.

7 MR. MATCH: Sorry.

8 THE COURT: No, you can sit down. The parties have
9 addressed all of the questions that I had. But let me give
10 each side an opportunity to add anything else they wanted to
11 add before we conclude the argument.

12 MS. HURST: Your Honor, briefly, I would just add on
13 behalf of Microsoft that this whole discussion is even more
14 theoretical because there's no theory in this complaint of how
15 there could be any infringement resulting from any act
16 allegedly taken by Microsoft. And that's important because
17 even for 1202(b)(1), it's not just removal in the abstract,
18 your Honor. It's removal with the ultimate aim of facilitating
19 some infringement. And putting aside the propriety of
20 analyzing scienter on a Rule 12(b) motion, your Honor, first
21 let's just go to the objective thing that is required, which is
22 some theory of infringement. And then there are these layers
23 of scienter around that, your Honor. But the Court doesn't
24 even have to look at those layers of scienter and whether the
25 allegations are sufficient, because there's no theory of

O63AIntO

1 ultimate infringement identified that is furthered by the
2 alleged acts. And for OpenAI's ChatGPT product, there's no
3 causation related -- between the removal of CMI for either
4 OpenAI or Microsoft, and any theory of infringement here, your
5 Honor.

6 So while we're skeptical that any amendment can cure
7 all of these problems, I did want to underscore, your Honor,
8 the degree of remoteness that is really presented here by the
9 allegations against Microsoft.

10 THE COURT: Thank you.

11 Anything further from plaintiff's counsel?

12 MR. MATCH: Yes, your Honor. I would just like to
13 touch on this issue as well. Our allegations of infringement
14 or infringement theory are contained in paragraphs 47 to 50.
15 And there's really a couple of different types of infringement.
16 So, first of all, it's not just inducing, enabling, or
17 facilitating infringement, which is one type of infringement --
18 or scienter.

19 There's also the concealment. So our theory, one of
20 our theories of infringement, is that they infringe by two
21 things. One is copying the articles in the first place and the
22 second is regurgitation. So they know that removing CMI, I
23 think, would conceal from ChatGPT users that they infringed
24 copyright by both copying and regurgitation.

25 Now, of course we don't allege actual dissemination of

O63AIntO

1 our works, but it's important to keep in mind and I think the
2 Court touched on this, we can try to do that if the Court
3 believes that that's required, but we don't need to actually
4 allege actual dissemination of our articles because concealment
5 only has to be likely from an ex ante perspective when they
6 remove the CMI. And we've alleged that it's likely here for a
7 couple of reasons.

8 One is we've alleged the prevalence of our works in
9 their training sets, there's thousands of them. One thing I
10 forgot to mention before is that if the Court is wondering
11 about how many articles are there, we put it in our brief.
12 OpenAI has admitted that in earlier versions of the training
13 set contains exactly 6,484 of our articles. So there is no
14 dispute, no reasonable dispute at least at the pleading stage
15 if we were to amend that our articles are in there. They've
16 admitted it.

17 The other is that according to a study that we cited
18 in the complaint, 45 percent of the ChatGPT responses contains
19 material identical to preexisting texts. So ChatGPT is a
20 serial plagiarizer and it has a lot of our works to plagiarize.
21 And I think those two facts put together make it likely that ex
22 ante that our works would be regurgitated in whole or in part,
23 and that would conceal from ChatGPT users --

24 THE COURT: Well, I will put the question that your
25 adversary in effect invited me to put. Have you tried yourself

O63AIntO

1 in preparation for this lawsuit to see if that's true?

2 MR. MATCH: No. Like I said, we didn't do that
3 deliberately because -- yeah.

4 THE COURT: Whether it was deliberate or not.

5 MR. MATCH: Sure. We didn't do it.

6 THE COURT: I just wanted to know whether you did it
7 or not.

8 MR. MATCH: No. So, or at least I don't believe we
9 did. And the other type of -- I'm sorry. I'm lost here.

10 THE COURT: No, no. How old are you?

11 MR. MATCH: Thirty-four.

12 THE COURT: You don't lose your memory until you're my
13 age. Now come on. Go ahead.

14 MR. MATCH: Actually, I should say I don't think that
15 we attempted to do that. There may have been like one or two
16 very small prompts that we did at a very early stage that we
17 didn't decide to pursue. I honestly don't fully remember. But
18 there were certainly no -- our data scientist who's the expert
19 never tried to do that.

20 THE COURT: Okay, well.

21 MR. MATCH: But I also wanted to point out that
22 defendants -- there's a second theory of infringement, which is
23 infringement by users. So if it's a likely ex ante that our
24 works would be regurgitated, whether or not they actually have
25 been, then they have reason to know that ChatGPT users would

O63AInto

1 further disseminate those works, and that's because the
2 defendants promote ChatGPT as a -- I'm sorry, as a tool to
3 generate content for a future audience.

4 And I think it's at least plausible that the existence
5 of CMI on a regurgitation, which wouldn't be there because the
6 CMI is removed from the training data, would at least be a
7 deterrent to future copying by ChatGPT users.

8 I think all the Court needs to assume for those
9 purposes is that users respect the law, either because they
10 have some inherent regard for it or because they don't want to
11 be liable for copyright infringement. That's all that I think
12 the Court needs to assume to find that theory of infringement
13 at least plausible enough. I would stress in discovery that
14 they have the prompts, I assume, in their databases, at least
15 for some users, and so that can be explored through discovery.

16 THE COURT: All right. I can see that defense counsel
17 is itching to say some final words, but he has the burden of
18 proof, so I will hear from defense counsel. I admire your
19 restraint in only indicating by your body language that you
20 wanted to be heard as opposed to standing up and saying I want
21 to be heard.

22 MR. GRATZ: Thank you, your Honor. I will be
23 extremely brief.

24 We heard that this might have happened to someone else
25 and that is not particularized harm. That is exactly what the

O63AIntO

1 situation in *TransUnion*. If someone else's content were
2 outputted, that does not create a harm and does not show a
3 likelihood of harm for the plaintiff.

4 The other thing that we heard that I want to just
5 respond to briefly is that their theory on the second scienter
6 element is that we should have known that this would conceal
7 infringement. And the question is just conceal what, from
8 whom, resulting in what harm to them? Right?

9 THE COURT: You're saying it involves many speculative
10 steps?

11 MR. GRATZ: And the biggest speculative step in the
12 middle is that their stuff comes out of ChatGPT. The issue we
13 have been talking about, and I think now we heard, well, they
14 tried a little, and it wasn't going well and --

15 THE COURT: I don't know we heard that. His memory
16 was uncertain on that.

17 MR. GRATZ: Yeah. Well, we may well find out.

18 THE COURT: All right. Although you are hoping you
19 won't because you're hoping you'll win the motion.

20 MR. GRATZ: Indeed, your Honor. We hope not to find
21 out.

22 THE COURT: All right. Well, I thank all counsel for
23 their excellent arguments.

24 As I'm sure you're all aware, a requirement for being
25 a federal judge is that you be technologically incompetent, and

O63AIntO

1 I am fully qualified. So it will take me a little while to get
2 through this. But I will certainly get you a bottom line order
3 at least by two weeks from now. Whether it will be a full
4 opinion, it might just be an order with leave to amend, under
5 many possibilities. I don't know. But at least you'll know
6 where things stand two weeks from today.

7 So, again, my thanks. This concludes the proceeding.

8 (Adjourned)